

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-153108

(43)Date of publication of application : 11.06.1996

(51)Int.Cl.

G06F 17/30

(21)Application number : 06-293149

(71)Applicant : CANON INC

(22)Date of filing : 28.11.1994

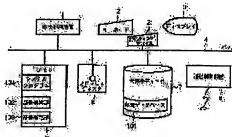
(72)Inventor : AIZAWA MICHIO
HIROTA MAKOTO
ITO SHIRO
SHIBATA SHOGO
UEDA TAKANARI
IKEDA YUJI
FUJITA MINORU

(54) METHOD AND DEVICE FOR RETRIEVING INFORMATION

(57)Abstract:

PURPOSE: To provide a method and a device for information retrieval which enables retrieval based upon a correct retrieval key word, specially, without depending upon the change of a retrieval key word unless the inputted retrieval key word is present, concretely, a method and device for information retrieval which can retrieve a department name after organization alteration even for a department name before organization alteration even when the department name to be retrieved is not present in a department data base owing to the alteration of the department name.

CONSTITUTION: When the inputted retrieval key word is not a retrieval key word of the data base, the retrieval key word of the data base which is closest to the inputted retrieval key word is estimated as the correct retrieval key word. In concrete, this device has the department data base 101 which stores a list of department names of the organization, a department retrieval part 102 which retrieves whether or not the department name to be retrieved is present in the department data base, and a department estimation part 103 which estimates the correct department name from the said department name to be retrieved unless the department name to be retrieved is present in the data base.



【特許請求の範囲】

【請求項1】 検索キーワードに基づいてデータベースの検索を行う情報検索方法において、

入力された検索キーワードが前記データベースの検索キーワードでない場合に、入力された前記検索キーワードに最も近い前記データベースの検索キーワードを、正しい検索キーワードとして推定することを特徴とする情報検索方法。

【請求項2】 前記データベースは、前記検索キーワードが変更された場合に、少なくとも部分的に変更前の検索キーワードに関連付けられた検索キーワードに変更されるデータベースであり、かつ古い検索キーワードを記憶しておかないデータベースであることを特徴とする請求項1記載の情報検索方法。

【請求項3】 検索キーワードが変更された場合に、新しい検索キーワードが少なくとも部分的に変更前の検索キーワードに関連付けられるデータベースであって、かつ古い検索キーワードを記憶しておかないデータベースと、

入力された検索キーワードでデータベースを検索する検索手段と、
前記入力された検索キーワードがデータベースに存在しない場合に、前記入力された検索キーワードに対して最も近い検索キーワードを正しい検索キーワードとして推定するキーワード推定手段とを備えることを特徴とする情報検索装置。

【請求項4】 前記キーワード推定手段は、
前記入力された検索キーワードを構成する文字に対して変更後の検索キーワードと一致する文字数を計算し、一致する文字数が最も多い検索キーワードを前記正しい検索キーワードとして推定する手段と、

前記入力された検索キーワードに対して形態素解析を行なって単語に分解し、変更後の検索キーワードと一致する単語数を計算して、その値が最大のものを前記正しい検索キーワードとして推定する手段と、

前記入力された検索キーワードの2文字以上の連続文字に対して変更後の検索キーワードと一致した連続数対応のポイントを与え、ポイント数が最大となるデータベース内の検索キーワードを前記正しい検索キーワードとして推定する手段と、

シソーラスなどの類義語の辞書を備え、前記入力されたキーワードに対する類義語に対して、データベースの検索キーワードと一致する単語数を計算し、一致する単語数が最も多いものを前記正しい検索キーワードとして推定する手段とを少なくとも1つの推定手段を備えることを特徴とする請求項3記載の情報検索装置。

【請求項5】 前記各推定手段において、前記入力されたキーワードから一般的共通な文字を比較対象より除外する除外手段を含むことを特徴とする請求項4記載の情報検索装置。

【請求項6】 組織の部署名の一覧を蓄えている部署データベースと、

検索対象の部署名が部署データベースに存在するかどうかを検索する部署検索手段と、

前記検索対象の部署名が部署データベースに存在しない場合に、前記検索対象の部署名から正しい部署名を推定する部署推定手段とを有することを特徴とする情報検索装置。

【請求項7】 前記部署推定手段は、前記検索対象の部署名を構成する文字に対して前記部署データベースの各部署名と一致する文字数を計算し、一致する文字数が最大となる部署データベースの中の部署名を正しい部署名として推定することを特徴とする請求項6記載の情報検索装置。

【請求項8】 前記部署推定手段は、前記検索対象の部署名に対して形態素解析を行なって単語に分割し、分割された各単語に対して前記部署データベースの各部署名と一致する単語数を計算し、一致する単語数が最大となる部署データベースの中の部署名を正しい部署名として推定することを特徴とする請求項6記載の情報検索装置。

【請求項9】 前記部署推定手段は、前記検索対象の部署名を構成する文字の2文字以上の連続文字に対して、部署データベースの各部署名と一致する連続文字を得て、連続文字数に対して与えられた単語に相応するポイント数を計算し、ポイント数が最大となる部署データベースの中の部署名を正しい部署名として推定することを特徴とする請求項6記載の情報検索装置。

【請求項10】 前記部署推定手段は、シソーラスなどの類義語の辞書を備え、前記検索対象の部署名に対する類義語に対して部署データベースの各部署名と一致する単語数を計算し、一致する単語数が最大となる部署データベースの中の部署名を正しい部署名として推定することを特徴とする請求項6記載の情報検索装置。

【請求項11】 前記部署推定手段は、前記検索対象の部署名の中から各部署名間でほぼ共通の文字を比較対象の部署名より除外することを特徴とする請求項7乃至10のいずれか1つに記載の情報検索装置。

【請求項12】 電子メールの配信に際して、前記推定した部署へ配信する配信手段を更に備えることを特徴とする請求項6記載の情報検索装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は入力された検索キーワードに基づいて情報検索を行う情報検索方法及びその装置に関するものである。

【0002】

【従来の技術】 従来の情報検索装置では、入力された検索キーワードに一致する文字列に基づき情報検索を行っている。

【0003】

【発明が解決しようとする課題】しかしながら、検索キーワードの入力ミスがあったり、検索キーワードが変更されて入力された検索キーワードがない場合には、希望の検索結果が得られず検索を繰り返す必要があった。このような例は検索キーワードに限らず、例えば、会社等で組織変更により部署名が代わった後に、組織変更前の部署名で送られてきたメール等に対して、対応する組織変更後の部署名を見つけてメールを送ること、あるいは組織変更前の部署名によるデータベースの検索はできなかった。

【0004】本発明は、前記従来の欠点を除去し、入力された検索キーワードが存在しない場合、特に検索キーワードの変更によらず正しい検索キーワードによる検索を可能とする情報検索方法及びその装置を提供する。また、具体的には、部署名の変更等により検索対象の部署名が部署データベースに存在しない場合にも、組織変更前の部署名に対して組織変更後の部署名を検索することができる情報検索方法及びその装置を提供する。

【0005】

【課題を解決するための手段】上述の課題を解決するために、本発明の情報検索方法は、検索キーワードに基づいてデータベースの検索を行う情報検索方法において、入力された検索キーワードが前記データベースの検索キーワードでない場合に、入力された前記検索キーワードに最も近い前記データベースの検索キーワードを、正しい検索キーワードとして推定することと特徴とする。ここで、前記データベースは、前記検索キーワードが変更された場合に、少なくとも部分的に変更前の検索キーワードに関連付けられた検索キーワードに変更されるデータベースであり、かつ旧い検索キーワードを記憶しておかないデータベースである。

【0006】又、本発明の情報検索装置は、検索キーワードが変更された場合に、新しい検索キーワードが少なくとも部分的に変更前の検索キーワードに関連付けられるデータベースであって、かつ旧い検索キーワードを記憶しておかないデータベースと、入力された検索キーワードでデータベースを検索する検索手段と、前記入力された検索キーワードがデータベースに存在しない場合に、前記入力された検索キーワードに対して最も近い検索キーワードを正しい検索キーワードとして推定するキーワード推定手段とを備える。ここで、前記キーワード推定手段は、前記入力された検索キーワードを構成する文字に対して変更後の検索キーワードと一致する文字数を計算し、一致する文字数が最も多い検索キーワードを前記正しい検索キーワードとして推定する手段と、前記入力された検索

キーワードの2文字以上の連続文字に対して変更後の検索キーワードと一致した連続数に対応のポイントを与え、ポイント数が最大となるデータベース内の検索キーワードを前記正しい検索キーワードとして推定する手段と、シソーラスなどの類義語の辞書を備え、前記入力されたキーワードに対する類義語に対して、データベースの検索キーワードと一致する単語数を計算し、一致する単語数が最も多いものを前記正しい検索キーワードとして推定する手段との少なくとも1つの推定手段を備える。また、前記各推定手段において、前記入力されたキーワードから一般的の共通文字を比較対象より除外する除外手段を含む。

【0007】又、本発明の情報検索装置は、組織の部署名の一覧を蓄えている部署データベースと、検索対象の部署名が部署データベースに存在するか否かを検索する部署検索手段と、前記検索対象の部署名が部署データベースに存在しない場合に、前記検索対象の部署名から正しい部署名を推定する部署推定手段とを有することと特徴とする。ここで、前記部署推定手段は、前記検索対象の部署名を構成する文字に対して前記部署データベースの各部署名と一致する文字数を計算し、一致する文字数が最大となる部署データベースの中の部署名を正しい部署名として推定する。また、前記部署推定手段は、前記検索対象の部署名に対して形態素解析を行なった単語に分割し、分割された各単語に対して前記部署データベースの各部署名と一致する単語数を計算し、一致する単語数が最大となる部署データベースの中の部署名を正しい部署名として推定する。また、前記部署推定手段は、前記検索対象の部署名を構成する文字の2文字以上の連続文字に対して、部署データベースの各部署名と一致する連続文字を得て、連続文字数に対して与えられた単語に相応するポイント数を計算し、ポイント数が最大となる部署データベースの中の部署名を正しい部署名として推定する。また、前記部署推定手段は、シソーラスなどの類義語の辞書を備え、前記検索対象の部署名に対する類義語に対して部署データベースの各部署名と一致する単語数を計算し、一致する単語数が最大となる部署データベースの中の部署名を正しい部署名として推定する。また、前記部署推定手段は、前記検索対象の部署名の中から各部署名間でほぼ共通の文字を比較対象の部署名より除外する。また、電子メールの配信に際して、前記推定した部署へ配信する配信手段を更に備える。

【0008】

【作用】以上の構成により、データベースに存在しない入力された検索キーワードに対して、近い検索キーワードを推定して検索することができる。

【0009】

【実施例】以下、本発明の実施例を添付図面を用いて詳細に説明する。

<実施例1> 図1は、本実施例の情報検索装置の一構成

例の図である。図中、1は装置の処理を行う中央処理装置（CPU）、2は操作入力のためのキーボード、21はマウス等の画面をポイントするためのポインティングデバイス、3は操作者への表示出力を行うディスプレイ、4はCPU1と他の構成要素を接続するためのバス、5はRAMやROMのメモリで、システムの処理を行うためのシステムプログラム104や、本実施例の構成要素である部署検索プログラムを含む部署検索部102と部署推定プログラムを含む部署推定部103を含む。

尚、これらのプログラムは、他の記憶部にあって処理の実行前にメモリ5へロードされてもよい。6はフロッピーディスクであり、必要なデータやプログラムの入力や出力のメディアとして用いる。7は大容量ディスクであり、ハードディスクや光磁気ディスク等の大容量記憶部であり、本実施例の構成要素である部署データベース101を含む。なお、部署データベース101は、フロッピーディスク6やメモリ5にあってよい。8は通信制御装置であり、通信回線を通じて本情報検索装置をアクセスするためのものである。

【0010】図2は、本実施例の情報検索装置の主要構成要素を示したものである。図2において、101は部署名の一覧を蓄えている上記部署データベースであり、102は検索対象の部署名がデータベースに存在するかどうか検索する上記部署検索部であり、103は部署データベースに存在しない誤った部署名に対し正しい部署名を推定する上記部署推定部である。

【0011】図3は、本実施例の情報検索装置の動作過程を示すフローチャートである。ステップS101で、操作者が検索したい部署名を入力する。ステップS102では、ステップS101で入力した部署名が部署データベース101に存在するかどうか部署検索部102で検索される。部署名がデータベース101に存在する場合はステップS103の処理へ進む。存在しない場合はステップS104の処理へ進む。ステップS103では、入力された部署名をこれから出力される部署名と定める。

【0012】ステップS102で部署名がデータベース101に存在しない場合は、ステップS104で、入力された部署データベース101に存在しない誤った部署名に対し、部署推定部103が近い部署名を推定する。ステップS105では、ステップS104で推定した部署名をこれから出力する部署名と定める。ステップS106では、ステップS103またはステップS105により得た部署名を出力する。

【0013】図4は、図3のステップS104に示す部署推定の処理をより詳しく説明するフローチャートである。ステップS201では、部署データベース101の各部署名について、入力された部署名と一致する文字数を計算し、ステップS202では、ステップS201で計算した文字数が最も多い部署名を、正しい部署名と推

定する。

【0014】以下、実際の例について詳しく説明する。図5に部署データベース101の組織変更前後の例を示す。この例では、部署データベース101で、組織変更により元の人工知能研究部、コンピュータ研究部、計算機画像研究部、特許部が、それぞれ新しく知能システム研究部、並列コンピュータ研究部、画像情報研究部、知的所有権保護部と改名されたことを表わしている。そこへ、組織変更後に「人工知能研究部」宛のメールが届いたとする。

【0015】この場合の処理の流れは以下のようになる。すなわち、図3のステップS101で、部署名が「人工知能研究部」と入力され設定される。次のステップS102で「人工知能研究部」が部署データベース101に存在しない誤った部署名であることが分かるので、ステップS104の処理へ進む。ステップS104では、まず図4に示すステップS201の一致文字数の計算を行う。計算した結果を図6に示す。ステップS202では、図6において一致する文字数が最も多い「知能システム研究部」を正しい部署名と推定する。このようにして、図3のステップS104の処理が終わる。

【0016】次のステップS105では、出力する部署名を「知能システム研究部」と定め、ステップ106で、「知能システム研究部」を部署名として出力する。従って、「人工知能研究部」宛のメールを「知能システム研究部」に送ることができる。なお、図3のステップS101は操作者でなく、通信回線からの入力を行うことができ、部署名の出力に代わって電子メールを自動送信することもできる。

【0017】尚、上記実施例では、部署名を構成する全ての文字に対して一致する文字数を計算していた。これに限らず、「部」とか「課」等の一般的な共通文字に対しては一致する文字数を計算しないようにしてもよい。又、上記実施例では、一致する文字数を計算しその値を用いて正しい部署名を推定していた。それに限らず、形態素解析を行ない単語に分割し、一致する単語数を計算しその値が最大のものを正しい部署名と推定してもよい。

【0018】又、2文字以上連続して一致する文字列の連続数に対して単語に相応するポイントを与えてこのポイントが最も大きいものを正しい部署名と推定してもよい。又、一致する単語だけでなく、シソーラスなどの情報から類義とわかる単語数も加える。例えば、「特許」に対し、類義である「知的所有権」も数に加えてもよい。

【0019】更に、上述した方法を適当に組み合わせてもよい。

<実施例2>以上の実施例では、会社組織等の部署名に対しての変更があった場合、メール等が変更前の部署名で送られてきた場合の情報検索装置の処理について説明

し、送られてきたメールに対して新しい部署名を検索することが可能となった。

【0020】従って、データベースを検索する情報検索装置では、本実施例の部署データベースのように部分的に名称の変更が起こる可能性があり、部署名に限らず旧名称を残さないデータベースに対して、本実施例で説明した情報検索装置がそのまま使用できることは言うまでもない。すなわち、検索のキーワードとなる部分が、部分的にまたはすべて前のキーワードに関連付けられたキーワードに変更が起こる可能性があるデータベースがあり、かつ旧キーワードとなる部分を記憶しておかない情報検索装置においては、前記データベースと、キーワードでデータベースを検索するキーワード検索部と、検索対象のキーワードがデータベースに存在しない場合、検索対象のキーワードに対して最も近いキーワードを推定するキーワード推定部を備えることで、検索が可能となる。

【0021】キーワードの推定部としては、上記実施例と同様に、キーワードを構成する全ての文字に対して一致する文字数を計算し、一致する文字数が最も多いキーワードを一致したキーワードを推定するか、キーワードに対して形態素解析を行ない単語に分解して一致する単語数を計算し、その値が最大のものをキーワードと推定するか、キーワードの2文字以上の連続文字に対して、キーワードと一致した連続数対応のポイントを予め与え、ポイント数が最大となるデータベース内のキーワードを一致したキーワードと推定するか、キーワード推定部にシソーラスなどの類義語の辞書を備え、検索するキーワードに対する類義語に対して、データベースの各キーワードの一致する単語数を計算し、一致する単語数が最も多いものをキーワードとして推定する。また、各キーワード間で一般的共通な文字をキーワードの比較対象より除外する。

【0022】以上のようにして、部分的にまたはすべて前のキーワードに関連付けられたキーワードに変更が起こる可能性があるデータベースであり、旧キーワードが残さないデータベースに対して、本発明の情報検索方法が普遍的に使用できる。尚、本発明は、複数の機器から構成されるシステムに適用しても、1つの機器から成る装置に適用しても良い。また、本発明はシステム或は装

置にプログラムを供給することによって達成される場合にも適用できることはいうまでもない。

【0023】

【発明の効果】以上説明したように、本発明は、入力された検索キーワードが存在しない場合、前記入力された検索キーワードから正しい検索キーワードを推定することにより、特に検索キーワードの変更によらず正しい検索キーワードによる検索を可能とした。

【0024】また、本発明は、検索対象の部署名が部署データベースに存在しない場合、誤った部署名に対し正しい部署名を推定する部署推定部を設けたことにより、組織変更前の部署名に対して組織変更後の部署名を検索することができるという効果がある。

【図面の簡単な説明】

【図1】実施例1を説明する為の情報検索装置の一構成例である。

【図2】本発明の論理構成を示す図である。

【図3】実施例の動作過程を示すフローチャートである。

【図4】実施例1の部署推定のステップ104をより詳しく説明する為のフローチャートである。

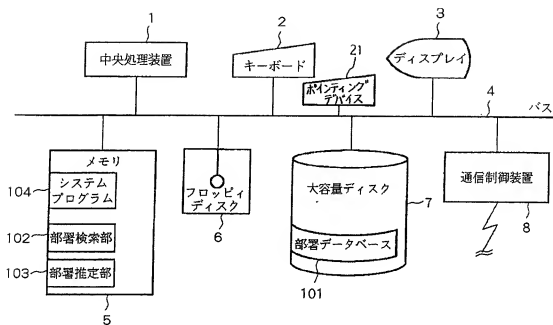
【図5】実施例1の部署データベース101の例を示す図である。

【図6】実施例1の部署推定ステップ104における、一致文字数の計算例を示す図である。

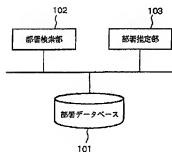
【符号の説明】

- 1 CPU
- 2 キーボード
- 3 ディスプレイ
- 4 バス
- 5 メモリ
- 6 フロッピーディスク
- 7 大容量ディスク
- 8 通信制御装置
- 21 ポインティングデバイス
- 101 部署データベース
- 102 部署検索部
- 103 部署推定部
- 104 システムプログラム

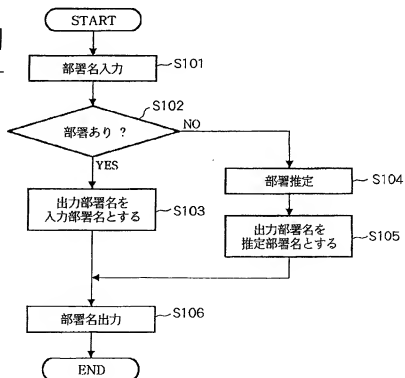
【図1】



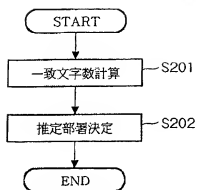
【図2】



【図3】



【図4】



【図5】

変更前の部署名	変更後の部署名 (部署データベース)
人工知能研究部	知能システム研究部
コンピュータ研究部	並列コンピュータ研究部
計算機画像研究部	画像情報研究部
特許部	知的所有権保護部

【図6】

部署名	一致文字数
知能システム研究部	5
並列コンピュータ研究部	3
画像情報研究部	3
知的所有権保護部	2

フロントページの続き

(72) 発明者 柴田 昇吾
東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内

(72) 発明者 上田 隆也
東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内

(72) 発明者 池田 裕治
東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内

(72) 発明者 藤田 稔
東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内

H08-153108

Machine Translation from Japanese Patent & Utility Model Gazette DB
(<http://www4.ipdl.inpit.go.jp/Tokujitu/tjsogodben.ipdl?N0000=115>)

CLAIMS

[Claim(s)]

[Claim 1] In an information retrieval method which searches a database based on a retrieval key word, An information retrieval method presuming a retrieval key word of said database nearest to [when an inputted retrieval key word is not a retrieval key word of said database] said inputted retrieval key word as a right retrieval key word.

[Claim 2] Said database is a database changed into a retrieval key word selectively related with a retrieval key word before change at least, when said retrieval key word is changed, And the information retrieval method according to claim 1 being a database which does not memorize an old retrieval key word.

[Claim 3] A database with which a new retrieval key word is selectively related with a retrieval key word before change at least when a retrieval key word characterized by comprising the following is changed.

And a database which does not memorize an old retrieval key word.

A search means to search a database with an inputted retrieval key word.

A keyword estimation means which presumes the nearest retrieval key word as a right retrieval key word to said inputted retrieval key word when said inputted retrieval key word does not exist in a database.

[Claim 4] The information retrieval device comprising according to claim 3:

A means for said keyword estimation means to calculate the number of characters which is in agreement with a retrieval key word after change to a character which constitutes said inputted retrieval key word, and to presume a retrieval key word with most characters in agreement as said right retrieval key word.

A means by which a morphological analysis is conducted to said inputted retrieval key word, it decomposes into a word, the number of words which is in agreement with a retrieval key word after change is calculated, and the value presumes the greatest thing as said right retrieval key word.

A means to presume a retrieval key word in a database which gives a point corresponding to a consecutive number which was in agreement with a retrieval key word after change to said inputted continuation character of two or more characters of a retrieval key word and with which a point size serves as the maximum as said right retrieval key word.

At least one estimation means with a means to have a dictionary of synonyms, such as a thesaurus, to calculate the number of words which is in agreement with a retrieval key word of a database to a synonym to said inputted keyword, and to presume what has most numbers of words in agreement as said right retrieval key word.

[Claim 5] The information retrieval device according to claim 4 by which an exclusion means to except a general common character from said inputted keyword from a comparison object being included in said each estimation means.

[Claim 6] An information retrieval device comprising:

A department database in which a list of names of department of an organization is stored.

A department search means to search whether a name of department of a retrieval object exists in a department database.

A department estimation means which presumes a right name of department from a name of department of said retrieval object when a name of department of said retrieval object does not exist in a department database.

[Claim 7]The information retrieval device according to claim 6 presuming a name of department in a department database with which said department estimation means calculates the number of characters which is in agreement with each post name of said department database to a character which constitutes a name of department of said retrieval object, and the number of characters in agreement serves as the maximum as a right name of department.

[Claim 8]Said department estimation means conducts a morphological analysis to a name of department of said retrieval object, and divides it into a word, The information retrieval device according to claim 6 presuming a name of department in a name-of-department database which calculates the number of words which is in agreement with each post name of said department database to each divided word, and with which the number of words in agreement serves as the maximum as a right name of department.

[Claim 9]As opposed to a continuation character of two-character ** of a character in which said department estimation means constitutes a name of department of said retrieval object, The information retrieval device according to claim 6 presuming a name of department in a department database which calculates a point size which ****s in a word given to the number of continuation characters by obtaining a continuation character which is in agreement with each post name of a department database, and with which a point size serves as the maximum as a right name of department.

[Claim 10]Said department estimation means is provided with a dictionary of synonyms, such as a thesaurus, and the number of words which is in agreement with each post name of a department database to a synonym to a name of department of said retrieval object is calculated, The information retrieval device according to claim 6 presuming a name of department in a department database with which the number of words in agreement serves as the maximum as a right name of department.

[Claim 11]An information retrieval device of any one statement of claim 7 thru/ or 10 which said department estimation means is between [out of a name of department of said retrieval object] each post names, and is characterized by excepting an almost common character from a name of department of a comparison object.

[Claim 12]The information retrieval device according to claim 6 having further a distribution means distributed to said presumed post on the occasion of distribution of an E-mail.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application]

This invention relates to the information retrieval method which performs information retrieval based on the inputted retrieval key word, and its device.

[0002]

[Description of the Prior Art] In the conventional information retrieval device, information retrieval is performed based on the character string which is in agreement with the inputted retrieval key word.

[0003]

[Problem(s) to be Solved by the Invention]

However, when there was an input mistake of a retrieval key word or there was no retrieval key word which the retrieval key word was changed and was inputted, desired search results were not obtained but there was ***** necessity about search. As opposed to the mail etc. which have been sent by the name of department before reorganization after a name of department replaces such an example by reorganization not only in a retrieval key word in a company etc., for example, Search of the database by the name of department before finding the name of department after corresponding reorganization, and sending e-mail or reorganization was not completed.

[0004] This invention provides the information retrieval method which is not based on change which is especially a retrieval key word, but enables search by a right retrieval key word, and its device, when said conventional fault is removed and the inputted retrieval key word does not exist. Also when the name of department of a retrieval object does not exist in a department database by change of a name of department, etc., specifically, the information retrieval method which can search the name of department after reorganization to the name of department before reorganization, and its device are provided.

[0005]

[Means for Solving the Problem]

In order to solve above-mentioned SUBJECT, an information retrieval method of this invention, In an information retrieval method which searches a database based on a retrieval key word, When an inputted retrieval key word is not a retrieval key word of said database, a retrieval key word of said database nearest to said inputted retrieval key word is presumed as a right retrieval key word. Here, when said retrieval key word is changed, said database is a database changed into a retrieval key word selectively related with a retrieval key word before change at least, and is a database which does not memorize an old retrieval key word.

[0006] An information retrieval device of this invention is provided with the following.

A database which a new retrieval key word is a database selectively related with a retrieval key word before change at least when a retrieval key word is changed, and does not memorize an old retrieval key word.

A search means to search a database with an inputted retrieval key word.

A keyword estimation means which presumes the nearest retrieval key word as a right retrieval key word to said inputted retrieval key word when said inputted retrieval key word does not exist in a database.

Said keyword estimation means calculates the number of characters which is in agreement with a retrieval key word after change to a character which constitutes said inputted retrieval key word here, A means to presume a retrieval key word with most characters in agreement as said right retrieval key word, Conduct a morphological analysis to said inputted retrieval key word, decompose into a word, and the number of words which is in agreement with a retrieval key word after change is calculated, A means by which the value presumes the greatest thing as said right retrieval key word, A point corresponding to a consecutive number which was in agreement with a retrieval key word after change to said inputted continuation character of two or more characters of a retrieval key word is given, A means to presume a retrieval key word in a database with which a point size serves as the maximum as said right retrieval key word, It has a dictionary of synonyms, such as a thesaurus, and has at least one estimation means with a means to calculate the number of words which is in agreement with a retrieval key word of a database, and to presume what has most numbers of words in agreement as said right retrieval key word, to a synonym to said inputted keyword. In said each estimation means, an exclusion means to except a general common character from said inputted keyword from a comparison object is included.

[0007] This invention is characterized by an information retrieval device comprising the following.

A department database in which a list of names of department of an organization is stored.

A department search means to search whether a name of department of a retrieval object exists in a department database.

A department estimation means which presumes a right name of department from a name of department of said retrieval object when a name of department of said retrieval object does not exist in a department database.

Here, said department estimation means presumes a name of department in a department database which calculates the number of characters which is in agreement with each post name of said department database to a character which constitutes a name of department of said retrieval object and with which the number of characters in agreement serves as the maximum as a right name of department. Said department estimation means conducts a morphological analysis to a name of department of said retrieval object, and divides it into a word, The number of words which is in agreement with each post name of said department database to each divided word is calculated, and the number of words in agreement presumes a name of department in a name-of-department database used as the maximum as a right name of department. Said department estimation means receives a continuation character of two-character ** of a character which constitutes a name of department of said retrieval object, A point size which ****s in a word given to the number of continuation characters is calculated by obtaining a continuation character which is in agreement with each post name of a department database, and a name of department in a department database with which a point size serves as the maximum is

presumed as a right name of department. Said department estimation means is provided with a dictionary of synonyms, such as a thesaurus, and presumes a name of department in a department database which calculates the number of words which is in agreement with each post name of a department database to a synonym to a name of department of said retrieval object and with which the number of words in agreement serves as the maximum as a right name of department. Said department estimation means is between [out of a name of department of said retrieval object] each post names, and excepts an almost common character from a name of department of a comparison object. On the occasion of distribution of an E-mail, it has further a distribution means distributed to said presumed post.

[0008]

[Function]

By the above composition, a near retrieval key word can be presumed and searched to the inputted retrieval key word which does not exist in a database.

[0009]

[Example]

Hereafter, the example of this invention is described in detail using an accompanying drawing.

<Example 1> drawing 1 is a figure of the example of 1 composition of the information retrieval device of this example. The central processing unit (CPU) with which one processes a device among a figure, and 2 The keyboard for an operational input, A bus for a display for pointing vice for 21 to point at screens, such as a mouse, and 3 to perform the display output to an operator and 4 to connect other components with CPU1 and 5 are the memories of RAM or ROM, The department retrieval part 102 containing the system program 104 for processing a system and the department retrieval program which is the components of this example, and the department estimating part 103 including a department presumption program are included. These programs are in other storage parts stores, and may be loaded to the memory 5 before execution of processing. 6 is a floppy disk and is used as media of required data, the input of a program, or an output. 7 is a mass disk, is mass memory units, such as a hard disk and a magneto-optical disc, and contains the department database 101 which is a component of this example. The department database 101 may be in the floppy disk 6 or the memory 5. 8 is a communication control unit and is a thing for accessing this information retrieval device via a communication line.

[0010]Drawing 2 shows the main components of the information retrieval device of this example. It is the above-mentioned department database in which 101 is storing the list of names of department in drawing 2, 102 is the above-mentioned department retrieval part with which the name of department of a retrieval object searches whether it exists in a database, and 103 is the above-mentioned department estimating part which presumes a right name of department to the name of department which did not make a mistake in existing in a department database.

[0011]Drawing 3 is a flow chart which shows the process of the information retrieval device of this example of operation. The name of department which an operator wants to search with Step S101 is inputted. It is searched with Step S102 by the department retrieval part 102 whether the name of department inputted at Step S101 exists in the department database 101. When a name of department exists in the database 101, it progresses to processing of Step S103. When it does not exist, it progresses to processing of Step S104. In Step S103, the inputted name of department is determined as the name of department outputted from now on.

[0012]When a name of department does not exist in the database 101 at Step S102, it is Step S104, and although inputted, the department estimating part 103 presumes a near name of department to the name of department which did not make a mistake in existing in the department database 101. In Step S105, it is determined as the name of department which outputs after this the name of department presumed at Step S104. In Step S106, the name of department obtained by Step S103 or Step S105 is outputted.

[0013]Drawing 4 is a flow chart which explains in more detail processing of department presumption shown in Step S104 of drawing 3. In Step S201, the number of characters which is in agreement with the inputted name of department about each post name of the department database 101 is calculated, and a name of department with most characters calculated at Step S201 is presumed to be a right name of department at Step S202.

[0014]Hereafter, it explains in more detail about a actual example. The example before and behind the reorganization of the department database 101 is shown in drawing 5. This example expresses that the original artificial intelligence research section, the computer research section, the computer picture research section, and the patent part were newly renamed by reorganization with the intelligence system research part, the parallel computer research section, the picture information research section, and the protection-of-intellectual-property part, respectively in the department database 101. Suppose there that the mail addressed to a "artificial intelligence research section" arrived after reorganization.

[0015]The flow of processing in this case is as follows. That is, at Step S101 of drawing 3, a name of department is inputted as a "artificial intelligence research section", and is set up. Since it turns out at the following step S102 that a "artificial intelligence research section" is the name of department which did not make a mistake in existing in the department database 101, it progresses to processing of Step S104. In Step S104, the number of coincidence characters of Step S201 first shown in drawing 4 is calculated. The calculated result is shown in drawing 6. In Step S202, an "intelligence system research part" with most characters which are in agreement in drawing 6 is presumed to be a right name of department. Thus, processing of Step S104 of drawing 3 finishes.

[0016]In the following step S105, the name of department to output is determined as an "intelligence system research part", and an "intelligence system research part" is outputted as a name of department at Step 106. Therefore, the mail addressed to a "artificial intelligence research section" can be sent to an "intelligence system research

part." Step S101 of drawing 3 can perform not an operator but the input from a communication line, and can also transmit an E-mail automatically instead of the output of a name of department.

[0017]In the above-mentioned example, the number of characters which is in agreement to all the characters which constitute a name of department was calculated. It may be made not to calculate the number of characters in agreement to general common characters, such as not only this but a "part", and a "section." In the above-mentioned example, the number of characters in agreement was calculated and the right name of department was presumed using the value. Not only it but a morphological analysis may be conducted, the number of words which divides into a word and is in agreement with it may be calculated, and the value may presume the greatest thing to be a right name of department.

[0018]The point which ****s in a word to the consecutive number of the character string which is continuously in agreement two or more characters may be given, and a thing with this largest point may be presumed to be a right name of department. Not only a word in agreement but the number of words which turns out to be class ** from information, including a thesaurus etc., is applied. For example, the "intellectual property rights" which is class ** may be added to a number to "a patent."

[0019]The method mentioned above may be combined suitably. In the example beyond <Example 2>, when there was change to names of department, such as a company organization, it became possible to explain and to search a new name of department to the sent mail about processing of the information retrieval device in the case of having been sent by the name of department before e-mail etc. changing.

[0020]Therefore, it cannot be overemphasized that the information retrieval device explained by this example can use it as it is to the database which change of a name may take place selectively like the department database of this example in the information retrieval device with which a database is searched, and does not leave not only a name of department but the old name. Namely, the keyword by which the portion used as the keyword of search was selectively related altogether with the front keyword has a database with which change may take place, And in the information retrieval device which does not memorize the portion used as the old keyword, Search becomes possible by having said database, a retrieval-by-keyword part which restrains a database by a keyword, and a keyword estimating part which presumes the nearest keyword to the keyword of a retrieval object when the keyword of a retrieval object does not exist in a database.

[0021]The number of characters which is in agreement like the above-mentioned example as an estimating part of a keyword to all the characters which constitute a keyword is calculated, . [whether the keyword which was in agreement in the keyword with most characters in agreement is presumed, and] Conduct a morphological analysis to a keyword, calculate the number of words which decomposes into a word and is in agreement with it, and the value presumes the greatest thing to be a keyword, or the

continuation character of two or more characters of a keyword is received, . [whether the keywords in the database which gives beforehand the point corresponding to the consecutive number which was in agreement with the keyword and with which a point size serves as the maximum are presumed to be the congruous keywords, and] The number of words which corresponds with a keyword estimating part to the synonym to the keyword which provides with and searches the dictionary of synonyms, such as a thesaurus, is calculated, and what has most numbers of words in agreement is presumed as a keyword. A general common character is excepted from the comparison object of a keyword between each keyword.

[0022]It is a database with which change may take place, and the information retrieval method of this invention can use it for the keyword altogether related with the front keyword selectively as mentioned above universally to the database which the old keyword does not leave. Even if it applies this invention to the system which comprises two or more apparatus, it may be applied to the device which comprises one apparatus. It cannot be overemphasized that this invention can be applied also when attained by supplying a program to a system or a device.

[0023]

[Effect of the Invention]

As explained above, when the inputted retrieval key word does not exist, by this invention, a right retrieval key word is presumed from said inputted retrieval key word. Therefore, it was not based in particular on change of the retrieval key word, but search by a right retrieval key word was enabled.

[0024]In this invention, when the name of department of a retrieval object did not exist in a department database, the department estimating part which presumes a right name of department to the mistaken name of department was provided. Therefore, it is effective in the ability to search the name of department after reorganization to the name of department before reorganization.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1]It is an example of 1 composition of the information retrieval device for describing Example 1.

[Drawing 2]It is a figure showing the composition with logic of this invention.

[Drawing 3]It is a flow chart which shows the process of an example of operation.

[Drawing 4]It is a flow chart for explaining Step 104 of department presumption of Example 1 in more detail.

[Drawing 5]It is a figure showing the example of the department database 101 of Example 1.

[Drawing 6]It is a figure in the department estimating step 104 of Example 1 showing the example computation of the number of coincidence characters.

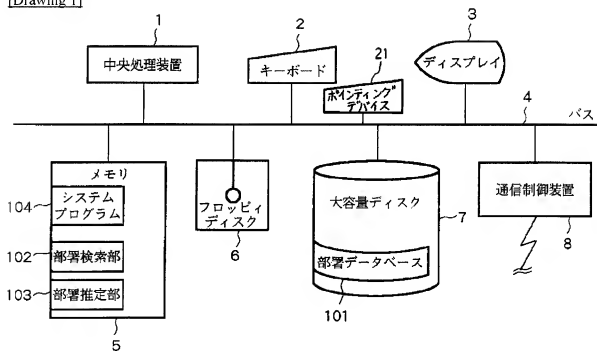
[Description of Notations]

1 CPU

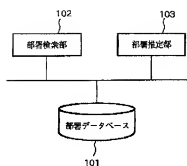
- 2 Keyboard
- 3 Display
- 4 Bus
- 5 Memory
- 6 Floppy disk
- 7 Mass disk
- 8 Communication control unit
- 21 Pointing device
- 101 Department database
- 102 Department retrieval part
- 103 Department estimating part
- 104 System program

DRAWINGS

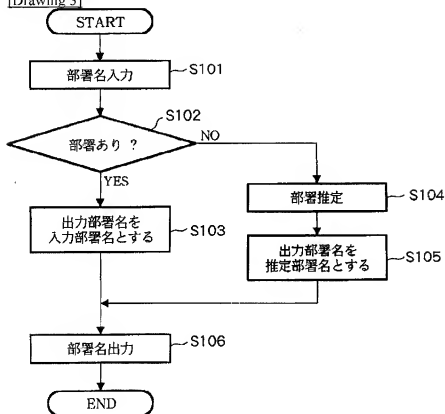
[Drawing 1]



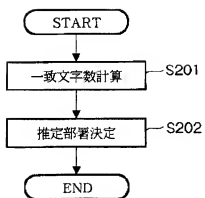
[Drawing 2]



[Drawing 3]



[Drawing 4]



[Drawing 5]

変更前の部署名	変更後の部署名 (部署データベース)
人工知能研究部	知能システム研究部
コンピュータ研究部	並列コンピュータ研究部
計算機画像研究部	画像情報研究部
特許部	知的所有権保護部

[Drawing 6]

部署名	一致文字数
知能システム研究部	5
並列コンピュータ研究部	3
画像情報研究部	3
知的所有権保護部	2